DOCUMENT PROCESSOR AND STORAGE MEDIUM

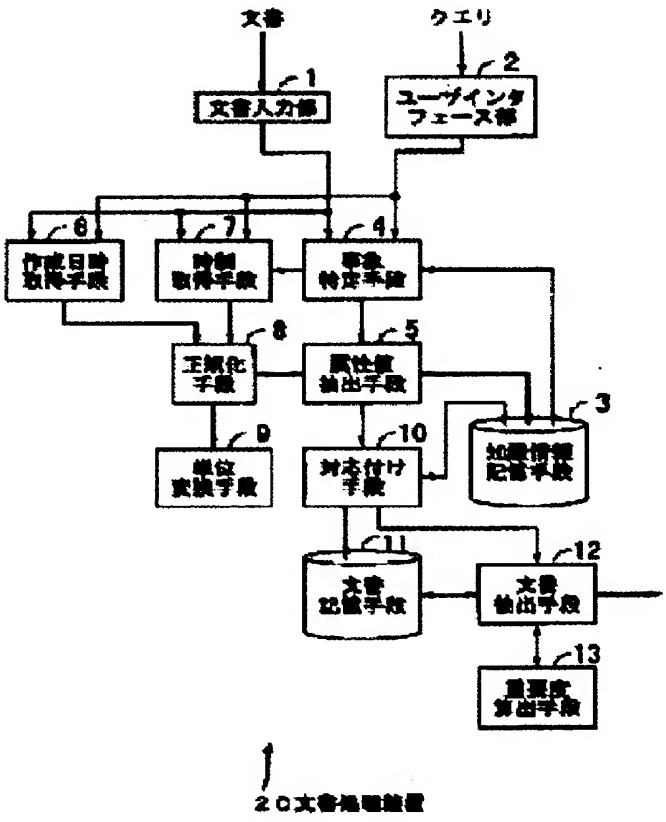
Publication number: JP11259498
Publication date: 1999-09-24
Inventor: HASHIMOTO MINAKO; KASHINO WAKAKO; OCHITANI AKIRA; NISHINO FUMITO
Applicant: FUJITSU LTD
Classification:
- international: ***G06F17/21; G06F17/30;*** **G06F17/21; G06F17/30;** (IPC1-7): G06F17/30; G06F17/21
- European: G06F17/30T1E
Application number: JP19980058384 19980310
Priority number(s): JP19980058384 19980310

Also published as:
WO9946698 (A1)
US6523025 (B1)
GB2350712 (A)
CN1286776 (A)
CN1109994C (C)

Report a data error here

Abstract of **JP11259498**

PROBLEM TO BE SOLVED: To enhance retrieving or clipping accuracy of a document. SOLUTION: The document to be processed is inputted from a document input part 1. A type of an event described on the inputted document is specified by referring to knowledge information stored in a knowledge information storage means 3 by an event specifying means 4. An attribute value of an attribute regarding the specified event is extracted from the document by an attribute value extracting means 5. A processing to make the attribute value extracted by the attribute value extracting means 5 correspond to substance in the real world is performed by a corresponding means 10. Information (normalized information) generated by the corresponding means 10 and information to specify the document or a place in which the document is stored are stored by correlating them by a document storage means 11. A query inputted from a user interface part 2 is compared with the normalized information and the information to specify the applicable document or the place in which the document is stored is outputted from the document storage means 11 by a document extracting means 12.



Data supplied from the ***esp@cenet*** database - Worldwide

(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（Ａ）　(11)特許出願公開番号

特開平11－259498

(43)公開日　平成11年(1999)９月24日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ０６Ｆ　17/30		Ｇ０６Ｆ　15/40	３７０Ａ
17/21		15/20	５９０Ｅ
		15/403	３１０Ａ
			３５０Ｃ

審査請求　未請求　請求項の数10　ＯＬ　（全 20 頁）

(21)出願番号　特願平10－58384

(22)出願日　平成10年(1998)３月10日

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号

(72)発明者　橋本　三奈子
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

(72)発明者　柏野　和佳子
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

(74)代理人　弁理士　服部　毅巖

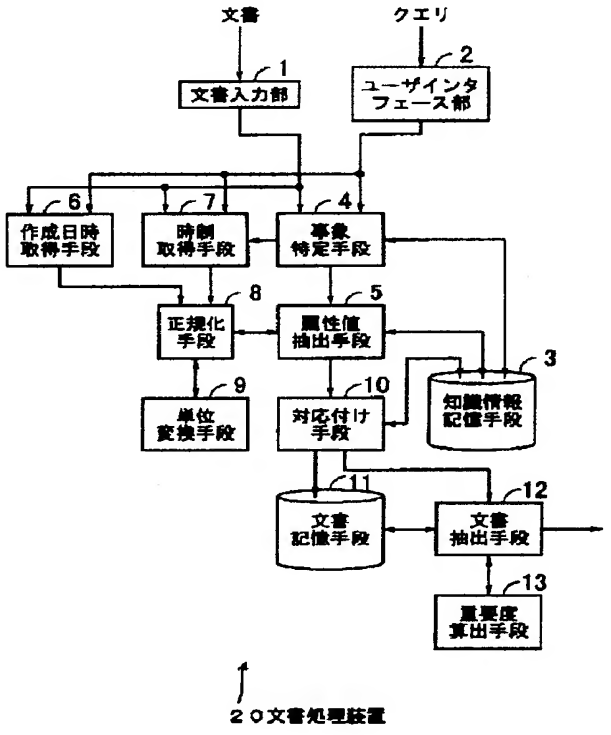
最終頁に続く

(54)【発明の名称】　文書処理装置および記録媒体

(57)【要約】

【課題】　文書の検索またはクリッピング精度を向上させる。

【解決手段】　文書入力部１からは処理の対象となる文書が入力される。事象特定手段４は、知識情報記憶手段３に記憶されている知識情報を参照して、入力された文書に記述されている事象の種類を特定する。属性値抽出手段５は、特定された事象に係わる属性の属性値を文書から抽出する。対応付け手段１０は、属性値抽出手段５によって抽出された属性値を、実世界の実体に対して対応付ける処理を行う。文書記憶手段１１は、対応付け手段１０によって生成された情報（正規化情報）と、文書または文書の格納場所を特定する情報とを関連付けて記憶する。文書抽出手段１２は、ユーザインタフェース部２から入力されたクエリと、正規化情報とを比較し、該当する文書または文書の格納場所を特定する情報を文書記憶手段１１から出力する。



1

【特許請求の範囲】

【請求項１】　入力された文書に対して所定の処理を施して記憶するとともに、与えられたクエリに対応する文書を、記憶されている文書の中から検索またはクリッピングする処理を行う文書処理装置において、

前記入力された文書を処理するために必要な知識情報を記憶する知識情報記憶手段と、

前記入力された文書に記述されている事象の種類を、前記知識情報記憶手段に記憶されている知識情報を参照して特定する事象特定手段と、

前記事象特定手段によって特定された事象に係わる属性の属性値を、前記知識情報記憶手段に記憶されている知識情報を参照して前記文書から抽出する属性値抽出手段と、

前記知識情報記憶手段に記憶された知識情報を参照して、前記属性値抽出手段によって抽出された属性値と、実世界の実体との対応付けを行う対応付け手段と、

前記対応付け手段によって対応付けがなされた属性値と、前記文書または前記文書の格納場所を特定するための情報とを関連付けて記憶する文書記憶手段と、

前記属性値と前記クエリとを参照して、対象となる文書に対して検索処理またはクリッピング処理を行う文書抽出手段と、

を有することを特徴とする文書処理装置。

【請求項２】　前記対応付け手段は、前記属性値の中で固有名であるものに対しては、他の属性値を参照してその固有名が示す実世界の実体を特定するとともに、特定された実体を一意に示す所定の情報を付与し、

前記文書抽出手段は、前記対応付け手段によって付与された前記所定の情報を参照して、検索処理またはクリッピング処理を行うことを特徴とする請求項１記載の文書処理装置。

【請求項３】　前記対応付け手段は、前記属性値が参照表現である「同」または「両」を含む場合に、それらの参照表現が参照する属性値を特定することを特徴とする請求項１記載の文書処理装置。

【請求項４】　前記文書に含まれているキーワードの出現頻度を参照して対象とする文書の重要度を算出する重要度算出手段を更に有し、

前記重要度算出手段は、前記対応付け手段によって参照先が特定された参照表現もキーワードと同等に処理することを特徴とする請求項３記載の文書処理装置。

【請求項５】　前記属性値のうち、数値化可能なものに関しては、対応する数値に変換して正規化する正規化手段を更に有し、

前記文書抽出手段は、前記正規化手段によって正規化された情報を参照して、検索処理またはクリッピング処理を行うことを特徴とする請求項１記載の文書処理装置。

【請求項６】　前記正規化手段によって得られた数値が有する単位を、他の所定の単位に変換する単位変換手段

2

を更に有することを特徴とする請求項５記載の文書処理装置。

【請求項７】　前記文書を構成する所定の文章の時制を取得する時制取得手段と、

前記文書が作成された日時を取得する作成日時取得手段とを更に有し、前記正規化手段は、前記時制取得手段によって取得された文書の時制と、前記作成日時取得手段によって取得された作成日時とを参照して、日時または期間を示す属性値の具体的な値を推定することを特徴とする請求項５記載の文書処理装置。

【請求項８】　前記文書に含まれているキーワードの出現頻度を参照して対象とする文書の重要度を算出する重要度算出手段を更に有し、

前記重要度算出手段は、前記正規化手段によって推定された日時または期間を加味して重要度を算出することを特徴とする請求項７記載の文書処理装置。

【請求項９】　前記事象特定手段、前記属性抽出手段、および、対応付け手段は、前記クエリに対しても文書の場合と同様の処理を行い、

前記文書抽出手段は、前記対応付け手段によって対応付けがなされた文書の属性値とクエリの属性値とを参照して、検索またはクリッピング処理を行うことを特徴とする請求項１記載の文書処理装置。

【請求項１０】入力された文書に対して所定の処理を施して記憶するとともに、与えられたクエリに対応する文書を、記憶されている文書の中から検索またはクリッピングする処理をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体において、

コンピュータを、

前記入力された文書を処理するために必要な知識情報を記憶する知識情報記憶手段、

前記入力された文書に記述されている事象の種類を、前記知識情報記憶手段に記憶されている知識情報を参照して特定する事象特定手段、

前記事象特定手段によって特定された事象に係わる属性の属性値を、前記知識情報記憶手段に記憶されている知識情報を参照して前記文書から抽出する属性値抽出手段、

前記知識情報記憶手段に記憶されている知識情報を参照して、前記属性値抽出手段によって抽出された属性値と、実世界の実体との対応付けを行う対応付け手段、

前記対応付け手段によって対応付けがなされた属性値と、前記文書または前記文書の格納場所を特定するための情報とを関連付けて記憶する文書記憶手段、

前記属性値と前記クエリとを参照して、対象となる文書に対して検索処理またはクリッピング処理を行う文書抽出手段、

として機能させるプログラムを記録したコンピュータ読[illegible]可能な記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は入力された文書に対して所定の処理を施して記憶するとともに、与えられたクエリに対応する文書を、記憶されている文書の中から検索またはクリッピングする処理を行う文書処理装置およびそのような処理をコンピュータに実行させるプログラムを記録した記録媒体に関する。

【０００２】

【従来の技術】近年、インターネットの普及や全文データベースの増加に伴って、個人が利用可能な情報が飛躍的に増加している。

【０００３】このような多量の情報の中から所望の情報を取得する方法としては、例えば、得ようとするデータの特徴を記述した検索式（クエリ）をキーとして、検索処理やクリッピング処理等を行う方法が一般的であった。

【０００４】

【発明が解決しようとする課題】しかし、従来の大規模な商用オンラインデータベースや全文検索システムでは、検索式の条件を緩やかにすると検索結果に含まれているノイズ（不要なデータ）が増加し、また、逆に厳しくすると検索洩れが発生するなど、ユーザが望む通りのデータを取得することが困難であるという問題があった。

【０００５】即ち、従来の文書フィルタリングにおける文書絞り込み処理や文書検索処理では、クエリと文書の内容の一致度や関連度に基づくランキング検索が行われている程度であるので、文書に含まれている重要情報やユーザの検索意図を十分に反映した文書絞り込みを行うことは困難であった。

【０００６】そのため、例えば、組織としての「橋本」が検索したいと思っても、「橋本」という地名が含まれた文書が検索されたりすることが多かった。また、２０万円台の新製品について検索したい場合には、「二十万円」「２０万円」、「二十一万円」、「二五万円」のように、あらゆる可能性を考慮して生成したクエリを用いる必要があった。

【０００７】更に、文書が作成された日時を用いて検索することは可能であっても、文書に含まれている日時情報を活用した検索を行うことができないという問題点があった。

【０００８】例えば、以下の文では、同じ「１日」でも示している日が異なる。

（ａ）Ａ社は、１日、Ｂ製品を発売する。

（ｂ）Ａ社は、１日、Ｂ製品を発売した。

【０００９】ここで、この文が作成された日が１９９７年２月１５日だとすると、（ａ）の場合では、「１日」は１９９７年３月１日を指し、また、（ｂ）では１９９７年２月１日を指すことになる。

【００１０】従来の方法においては、文書中の日時情報の属性を認知した上で、このような情報を検索に使用（活用）することが困難であるという問題点があった。本発明はこのような点に鑑みてなされたものであり、ユーザの検索意図を十分に反映した文書検索または文書絞り込みを行うことが可能な文書処理装置を提供することを目的とする。

【００１１】また、本発明は、ユーザの検索意図を十分に反映した文書検索処理またはクリッピング処理を行う文書処理を行うプログラムを記録した記録媒体を提供することを目的とする。

【００１２】

【課題を解決するための手段】図１は、上記目的を達成する本発明の原理図である。本発明は、入力された文書に対して所定の処理を施して記憶するとともに、与えられたクエリに対応する文書を、記憶されている文書の中から検索またはクリッピングする処理を行う文書処理装置において、知識情報記憶手段３と、事象特定手段４と、属性値抽出手段５と、対応付け手段１０と、文書記憶手段１１と、文書抽出手段１２とから構成されている。

【００１３】ここで、知識情報記憶手段３は、入力された文書を処理するために必要な知識情報を記憶する。事象特定手段４は、入力された文書に記述されている事象の種類を、知識情報記憶手段３に記憶されている知識情報を参照して特定する。属性値抽出手段５は、事象特定手段４によって特定された事象に係わる属性の属性値を、知識情報記憶手段３に記憶されている知識情報を参照して文書から抽出する。対応付け手段１０は、知識情報記憶手段３に記憶されている知識情報を参照して、属性値抽出手段５によって抽出された属性値と、実世界の実体との対応付けを行う。文書記憶手段１１は、対応付け手段１０によって対応付けがなされた属性値と、文書または文書の格納場所を特定するための情報とを関連付けて記憶する。文書抽出手段１２は、属性値とクエリとを参照して、対象となる文書に対して検索処理またはクリッピング処理を行う。

【００１４】知識情報記憶手段３には、事象とそれに関する属性、および、属性を構成する属性値を抽出するための情報とが関連付けられて記憶されている。事象特定手段４は、入力された文書と、知識情報記憶手段３に記憶されている知識情報とを照合することにより、文書中に記述されている事象を特定する。属性値抽出手段５は、知識情報記憶手段３を参照して、特定された事象に関する属性の属性値を文書から抽出する。対応付け手段１０は、知識情報記憶手段３に記憶されている知識情報を参照して、抽出された属性値と実世界の実体とを１対１に対応付ける処理を行う。文書記憶手段１１は、対応付けがなされた属性値と文書または文書の格納場所を特定するための情報とを関連付けて記憶する。文書抽出手

段１２は、入力されたクエリに含まれている情報と、文書記憶手段１１に記憶されている属性値とを照合することにより所望の文書を抽出する。

【００１５】これにより、対象となる文書に記述されている内容を、事象という観点から把握し、把握した事象を構成する属性の属性値を抽出するとともに、抽出した属性値を実世界の実体と対応付けることによって生成された情報を参照して文書を検索またはクリッピングするようにしたので、検索またはクリッピングの精度を向上させることが可能となる。

【００１６】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照して説明する。図１は、本発明の実施の形態の構成例を示すブロック図である。この図において、文書入力部１からは、処理対象の文書が入力される。ユーザインタフェース部２は、ユーザからのクエリを受け付ける。

【００１７】知識情報記憶手段３は、後述する事象やその事象に係わる属性に関する情報および固有名を固有コードに変換するための情報を記憶している。事象特定手段４は、知識情報記憶手段３に記憶されている知識情報（事象の種類に関する情報）を参照して、文書入力部１またはユーザインタフェース部２から入力された文書やクエリに記述されている事象の種類を特定する。

【００１８】ここで、「事象」とは、実世界で生起する「できごと」を示している。例えば、新聞記事などでは「Ａ社がＸを発売する。」というような実世界で発生した（または、これから発生する）事象に種々の補足情報が付加されて記述されていると考える。

【００１９】従って、事象特定手段４に対して、例えば、前述の「Ａ社がＸを発売する。」が入力されると、この文章に記述されている事象は、＜新製品の発売＞であると特定されることになる。ここで、＜＞は、その内部の語句が抽象化されて得られた概念であることを示している。

【００２０】なお、新聞記事などのように、記述の対象となる事象が明確であり、また、その表現様式が限られている文書においては、記述された事象のとりうる構造（以下、事象構造と適宜略記する）にも所定の制約条件が自ずと生ずることから、このような「事象」に着目して文書を解析することにより、効果的な処理を行うことが可能となる。

【００２１】属性値抽出手段５は、知識情報記憶手段３に記憶されている知識情報（所定の事象に係わる属性に関する情報）を参照して、文書またはクエリから属性値を抽出する。

【００２２】例えば、知識情報記憶手段３は、前述の＜新製品の発売＞という事象に関しては、＜販売会社＞、＜製品情報＞、＜日付＞、および、＜変更点＞などの属性を記憶しており、属性値抽出手段５は、事象特定手段４によって特定された事象に対応する属性を知識情報記憶手段３から取得し、その属性に対応する属性値を文書またはクエリから抽出する。

【００２３】例えば、前述の「Ａ社がＸを発売する。」という事象では、属性＜販売会社＞に対応する属性値「Ａ社」が取得され、また、属性＜製品情報＞に対応する属性値「Ｘ」などが抽出される。

【００２４】作成日時取得手段６は、文書またはクエリの作成日時を取得する。時制取得手段７は、文書またはクエリを構成する文章の時制（tense ）を取得する。正規化手段８は、属性値抽出手段５によって抽出された属性のうち、数値に変換可能なものを選択し、対応する数値に変換（正規化）する。

【００２５】単位変換手段９は、正規化手段８が正規化した数値の単位を変換する処理を行う。対応付け手段１０は、知識情報記憶手段３に記憶されている知識情報を参照して、属性値抽出手段５によって抽出された属性値を、実世界における実体に対応付ける処理を行う。なお、この「実体」とは、文書に記述されている属性値が示す実世界の「オブジェクト」を意味している。例えば、前述の例では、「Ａ社」と呼ばれる企業が複数存在する場合には、文書中に記述されている「Ａ社」がどの企業（オブジェクト）を示しているのかを特定する必要があるので、対応付け手段１０は、文書中の他の属性値（例えば、「社長の名前」や「本社地」など）を参照して「Ａ社」を特定する。

【００２６】文書記憶手段１１は、対応付け手段１０によって対応付けがなされた属性値集合と、もとの文書（または、もとの文書の格納場所を特定する情報）とを対応付けて記憶する。

【００２７】文書抽出手段１２は、対応付け手段１０から供給されたクエリに対応する文書を、属性値を参照して文書記憶手段１１から取得する。そして、重要度算出手段１３によって算出された個々の文書の重要度を参照し、ある閾値以上の重要度を有する文書を出力する。

【００２８】重要度算出手段１３は、所定のキーワードの出現頻度などを求めることにより、対象とする文書の重要度を算出する。図２を参照して、図１に示す実施の形態を含む通信システムの構成例について説明する。

【００２９】図２において、図１に示す文書処理装置２０は、例えば、インターネットなどのネットワーク２１に接続されている。このネットワーク２１には、端末装置２２ａ，２２ｂや、サーバ２３などが接続されている。

【００３０】端末装置２２ａ，２２ｂは、その入力部をユーザが操作して入力したクエリなどを受け付けて、文書処理装置２０に送信するとともに、送信したクエリに対応する文書を文書処理装置２０が送信した場合には、その文書を受信して、例えば、ＣＲＴ（Cathode Ray Tube）モニタなどへ表示出力する。

【００３１】サーバ２３は、記憶部２３ａに記憶している文書や画像などの情報を、ネットワーク２１を介して、要求を行った装置に対して送信する。文書処理装置２０は、端末装置２２ａ，２２ｂなどから送信されたクエリを記憶しておき、例えば、サーバ２３から新たな文書が供給された場合には、その文書と前述のクエリとの関連度が高い場合には、端末装置２２ａまたは端末装置２２ｂに対して送信する。

【００３２】次に、以上の実施の形態の動作について説明する。図３は、図１に示す実施の形態において、新たな文書が文書入力部１から入力された場合（例えば、図２に示すサーバ２３から新たな文書が供給された場合）に実行される処理の一例を説明するフローチャートである。

【００３３】このフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ１］文書入力部１は、新たな文書を入力する。

［Ｓ２］事象特定手段４は、文書に記述されている事象の種類を特定する。

【００３４】即ち、事象特定手段４は、知識情報記憶手段３に記憶されている事象と表現とのマッピング規則情報（図５参照）を参照して、文書に記述されている事象の種類を特定する。図５に示すマッピング規則では、“module”“end ”で囲繞された部分が一つの事象（または実体）と表現とのマッピング規則であり、１つの事象に対する表現のバリエーションを記述している。なお、図５に示すマッピング規則の詳細については後述する。

［Ｓ３］属性値抽出手段５は、知識情報記憶手段３に記憶されている知識情報を参照して属性値を抽出する。

【００３５】例えば、属性値抽出手段５は、図５に示す「module main 」内に記述されている事象のバリエーションのうち、入力された文書に対応する定義に含まれている属性（例えば、＜会社情報＞、＜製品＞等）の属性値を、他の「module」や「def 」などを参照することにより文書から抽出する。例えば、属性＜会社情報＞に対応する属性値は、第１７行目～第２１行目に記述されている定義および、第１２行目～第１５行目に記述されている定義を参照してパターンマッチングを行うことにより、文書から抽出する。

［Ｓ４］正規化手段８は、抽出された属性値に日付表現が含まれているか否かを判定し、日付表現が含まれている場合にはステップＳ５に進み、それ以外の場合にはステップＳ７に進む。

［Ｓ５］作成日時取得手段６は、文書が作成された日時を取得し、また、時制取得手段７は、対象となる事象が記述されている文章の時制（tense ）を取得する。

［Ｓ６］正規化手段８は、取得した文書作成日時情報と時制情報とを参照して、日付表現を対応する数値に変換する「日付表現変換処理」を実行する。

【００３６】なお、この処理の詳細については、図６を参照して後述する。

［Ｓ７］正規化手段８は、抽出された属性値に金額表現が含まれているか否かを判定し、金額表現が含まれている場合には、ステップＳ８に進み、それ以外の場合にはステップＳ１１に進む。

［Ｓ８］正規化手段８は、注目している金額表現が規定された通貨単位であるか否かを判定し、規定された通貨単位である場合にはステップＳ１０に進み、それ以外の場合にはステップＳ９に進む。

【００３７】例えば、規定されている通貨単位が「円」ある場合に、「＄」を単位とする金額表現が存在している場合にはステップＳ９に進む。

［Ｓ９］単位変換手段９は、内蔵されている記憶部に記憶している交換レートを読み出し、規定されている通貨単位に変換する処理を行う。

【００３８】例えば、「＄１００」という表現が存在している場合に、交換レートが「１＄＝１３０円」であるとすると、「＄１００」は「１３０００円」に変換されることになる。

［Ｓ１０］正規化手段８は、金額表現を数値に変換する「金額表現変換処理」を行う。なお、この処理の詳細は、図１３を参照して詳述する。

【００３９】前述の例では、「１３０００円」（文字列）が「13000 」（数値）に変換される。

［Ｓ１１］正規化手段８は、他の数値表現が存在するか否かを判定し、他の数値表現が存在する場合にはステップＳ１２に進み、それ以外の場合にはステップＳ１３に進む。

【００４０】例えば、「出荷台数５００００台」などの表現が存在する場合には、ステップＳ１２に進む。

［Ｓ１２］正規化手段８は、属性値に含まれている数値表現を対応する数値に変換する。例えば、前述の例では、「５００００」（文字列）が計算可能な数値「5000」に変換されることになる。

［Ｓ１３］対応付け手段１０は、属性値に固有名（例えば、「橋本電気」等）が含まれているか否かを判定し、含まれている場合にはステップＳ１４に進み、それ以外の場合にはステップＳ１５に進む。

［Ｓ１４］対応付け手段１０は、固有名を抽出し、その固有名に対応する固有名コードを知識情報記憶手段３から取得して付与する。

【００４１】例えば、前述の「橋本電気」に対応する固有名コード「００００１１」が、知識情報記憶手段３から読み出されて付与される。なお、知識情報記憶手段３には、関連する固有名を相互に関連づけて生成した情報が記憶されているので、文書中のある固有名が複数の候補を有する場合には、関連付けられている他の固有名を参照することにより、固有名を正確に特定することが可能となる。

【００４２】即ち、「橋本電気」に対応する候補が「橋本電気株式会社」と「株式会社橋本電気」の２つである場合（同名の会社が存在する場合）には、文書中に記載されている、例えば、社長名や所在地などと、知識情報記憶手段３に関連付けられて記憶されている固有名とを比較することにより、これらの候補を絞り込んで正確な固有名を取得することができる。

［Ｓ１５］対応付け手段１０は、参照表現（「同」または「両」などの表現）が存在するか否かを判定し、参照表現が存在する場合にはステップＳ１６に進み、それ以外の場合にはステップＳ１８に進む。

【００４３】例えば、参照表現である（同）が存在する場合には、ステップＳ１６に進むことになる。

［Ｓ１６］対応付け手段１０は、参照表現が参照している対象を同定する。

【００４４】例えば、「橋本電気（中山社長）は、橋本電算機（同）の独立を・・・」の場合では、参照表現「同」が参照している対象として「中山社長」を同定する。なお、この同定方法としては、「同」や「両」が内挿された括弧“（）”が検出された場合には、その括弧より前に出現する括弧内の属性値を、参照表現が参照している対象であると同定するようにすればよい。

［Ｓ１７］対応付け手段１０は、参照表現が参照している対象に対応する固有コードを取得し、取得した固有名コードを参照表現に対して付与する。

【００４５】前述の例では、「中山社長」の固有コード「０００１」が参照表現「同」に付与されることになる。

［Ｓ１８］対応付け手段１０は、正規化された属性値（以下、正規化情報と略記する）と、元の文書（または、元の文書の格納場所を特定するための情報）とを関連付けて文書記憶手段１１に記憶させる。

【００４６】以上の処理により、入力された文書に記述されている事象が特定されるとともに、その事象に係わる属性の属性値が取得される。そして、取得された属性値と実世界の実体とが対応付けられて得られた正規化情報と、もとの文書（または、元の文書の格納場所を特定するための情報）とが文書記憶手段１１に記憶されることになる。

【００４７】次に、図３のステップＳ６に示す、「日付表現変換処理」の詳細について説明する。図６は、図３に示す「日付表現変換処理」の詳細を説明するフローチャートである。このフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ３０］作成日時取得手段６は、文書の作成日時を取得して、%docyear, %docmonth, %docday に代入する。なお、文書の作成日時は、例えば、新聞記事であればその記事の発行日を取得する。また、新聞記事以外の文書であれば、ファイルの属性を参照して、その作成日時を取得する。

［Ｓ３１］正規化手段８は、属性値から日付表現を抽出する。

【００４８】例えば、対象となる文章が「橋本電気は新たなコンピュータを１日に発売。」であれば、日付表現として「１日」を抽出する。

［Ｓ３２］正規化手段８は、抽出した日付表現が数字と「年」、「月」、または、「日」によって構成されているか否かを判定し、ＹＥＳと判定した場合にはステップＳ３３に進み、それ以外の場合にはステップＳ３４に進む。

【００４９】例えば、前述の「１日」の例では、数字「１」と「日」とによって構成されていることから、ＹＥＳと判定されてステップＳ３３に進む。

［Ｓ３３］正規化手段８は、数字変換テーブル（図７参照）を参照して、日付表現を変換する処理を行う。

【００５０】図７に示す数字変換テーブルでは、数字表現に対応する正規化数値が関連付けられており、ある数字表現（文字列）が与えられると、その表現に対応する数値が与えられることになる。

［Ｓ３４］正規化手段８は、図８に示す日時表現変換テーブルを参照して、日時表現を対応する数値に変換する。

【００５１】図８に示す日時表現変換テーブルでは、表現と、そのタイプと、対応する正規化数値とが対応付けられている。なお、タイプとは表現の型であり、例えば、「ｄａｔｅ」は特定の日時を示し、また、「ｄａｔｅｒａｎｇｅ」は所定の期間を示している。例えば、１９９８年に作成された文書に「昨年の３月４日」という表現があれば、%year には（%docyear- １）＝（1998-1）=1997 が代入され、また、%monthと%dayには、それぞれ“３”と“４”が代入される。

【００５２】また、１９９７に作成された文書に「１９９８年の春」という表現があれば、%year には“1998”が代入されるので、from 1998-3-1 to 1998-5-30という正規化値が得られることになる。

【００５３】なお、この日時表現テーブルは、一例であるので、図示したテーブル以外にも種々の実施の形態が考えられる。

［Ｓ３５］正規化手段８は、全ての値が取得されたか否かを判定し、取得されたと判定した場合にはステップＳ３７に進み、それ以外の場合にはステップＳ３６に進む。

【００５４】例えば、年月日に対応する全ての値が取得された場合にはステップＳ３７に進む。

［Ｓ３６］正規化手段８は、日付の推定処理を行う。なお、この処理の詳細は、図９を参照して後述する。

［Ｓ３７］正規化手段８は、正規化された数値を%year, %month, %dayにそれぞれ代入して処理を終了する。

【００５５】以上の処理によれば、文書に含まれている日付表現を、対応する数値に変換することが可能とな

る。次に、図９を参照して図６のステップＳ３６に示す「日付推定処理」の詳細について説明する。図９に示すフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ５０］正規化手段８は、%year のみ値が未代入であるか否かを判定し、未代入である場合にはステップＳ５１に進み、それ以外の場合にはステップＳ５２に進む。

［Ｓ５１］正規化手段８は、%year 推定処理を行う。なお、この処理の詳細については、図１０を参照して後述する。

［Ｓ５２］正規化手段８は、%day以外の値が未代入であるか否かを判定し、未代入である場合にはステップＳ５３に進み、それ以外の場合にはステップＳ５５に進む。

［Ｓ５３］正規化手段８は、%month推定処理を行う。なお、この処理の詳細については、図１１を参照して詳述する。

［Ｓ５４］正規化手段８は、%year 推定処理を行う。

［Ｓ５５］正規化手段８は、%month以外の値が未代入であるか否かを判定し、その結果、未代入である場合にはステップＳ５６に進み、それ以外の場合にはステップＳ５８に進む。

［Ｓ５６］正規化手段８は、%day推定処理を行う。なお、この処理の詳細については図１２を参照して詳述する。

［Ｓ５７］正規化手段８は、%year 推定処理を行う。

［Ｓ５８］正規化手段８は、%year 以外の値が未代入であるか否かを判定し、未代入の場合にはステップＳ５９に進み、それ以外の場合には処理を終了する。

［Ｓ５９］正規化手段８は、推定日時を「from %year-1-1 to %year-12-31 」とする。即ち、%year以外の値が未代入である場合には、検索漏れが生ずることを防止するために、正規化値をできるだけ広い値に設定する。

【００５６】次に、図１０を参照して、図８のステップＳ５１，Ｓ５４，Ｓ５７に示す「%year 推定処理」の詳細について説明する。このフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ６０］正規化手段８は、対象となる文章から時制取得手段７によって取得された時制が未来形である場合にはステップＳ６１に進み、それ以外の場合にはステップＳ６５に進む。

［Ｓ６１］正規化手段８は、%docmonth が%monthよりも大きいか否かを判定し、大きいと判定した場合にはステップＳ６２に進み、それ以外の場合にはステップＳ６３に進む。

［Ｓ６２］正規化手段８は、%year に値（%docyear+1）を代入する。

【００５７】例えば、文書が作成された月が４月である場合に、「３月に・・・する予定である」という表現が文章中にある場合には、この「３月」は、来年の３月を示していると推定されることから、%year には値（%docyear+1）が代入される。

［Ｓ６３］正規化手段８は、%docmonth が%month以下であるか否かを判定し、ＹＥＳと判定した場合にはステップＳ６４に進み、それ以外の場合にはステップＳ６５に進む。

［Ｓ６４］正規化手段８は、%year に%docyearの値を代入する。

［Ｓ６５］正規化手段８は、時制取得手段７によって取得された時制が過去であるか否かを判定し、ＹＥＳと判定した場合にはステップＳ６６に進み、それ以外の場合には図９の処理に復帰（リターン）する。

［Ｓ６６］正規化手段８は、%docmonth の値が%monthの値以上であるか否かを判定し、ＹＥＳと判定した場合にはステップＳ６７に進み、それ以外の場合にはステップＳ６８に進む。

［Ｓ６７］正規化手段８は、%docyearの値を%year に代入する。

［Ｓ６８］正規化手段８は、%docmonth の値が%monthの値よりも小さいか否かを判定し、ＹＥＳと判定した場合にはステップＳ６９に進み、それ以外の場合には図９の処理に復帰する。

［Ｓ６９］正規化手段８は、値（%docyear-1）を%year に代入する。

【００５８】例えば、文書が作成された月が４月である場合に、「６月に・・・した」という表現が文章中にある場合には、この「６月」は、昨年の６月を示していると推定されることから、%year には値（%docyear-1）が代入される。

【００５９】次に、図１１を参照して、図９のステップＳ５３に示す「%month推定処理」の詳細について説明する。このフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ８０］正規化手段８は、時制取得手段７によって取得された対象となる文章の時制が未来形である場合にはステップＳ８１に進み、それ以外の場合にはステップＳ８５に進む。

［Ｓ８１］正規化手段８は、%docday が%dayよりも大きいか否かを判定し、ＹＥＳと判定した場合にはステップＳ８２に進み、それ以外の場合にはステップＳ８３に進む。

［Ｓ８２］正規化手段８は、%monthに値（%docmonth+1）を代入する。

【００６０】例えば、文書が作成された日が２日である場合に、「４日に・・・する予定である」という表現が文章中にある場合には、この「４日」は、次の月の４日を示していると推定されることから、%monthには値（%docmonth+1）が代入されることになる。

［Ｓ８３］正規化手段８は、%docday の値が%dayの値以下であるか否かを判定し、ＹＥＳと判定した場合にはステップＳ８４に進み、それ以外の場合にはステップＳ８

５に進む。

［Ｓ８４］正規化手段８は、%monthに%docmonth の値を代入する。

［Ｓ８５］正規化手段８は、時制取得手段７によって取得された時制が過去であるか否かを判定し、ＹＥＳと判定した場合にはステップＳ８６に進み、それ以外の場合には図９の処理に復帰（リターン）する。

［Ｓ８６］正規化手段８は、%docday の値が%dayの値以上であるか否かを判定し、ＹＥＳと判定した場合にはステップＳ８７に進み、それ以外の場合にはステップＳ８８に進む。

［Ｓ８７］正規化手段８は、%docmonth の値を%monthに代入する。

［Ｓ８８］正規化手段８は、%docday の値が%dayの値よりも小さいか否かを判定し、ＹＥＳと判定した場合にはステップＳ８９に進み、それ以外の場合には図９の処理に復帰する。

［Ｓ８９］正規化手段８は、値（%docmonth-1 ）を%monthに代入する。

【００６１】例えば、文書が作成された日が４日である場合に、「６日に・・・した」という表現が文章中にある場合には、この「６日」は、前の月の６日を示していると推定されることから、%monthには値（%docmonth-1 ）が代入される。

【００６２】次に、図１２を参照して、図９のステップＳ５６に示す「%day推定処理」の詳細について説明する。このフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ１００］正規化手段８は、%monthの値が１，３，５，６，８，１０，または，１２のうちの何れかであるか否かを判定し、ＹＥＳと判定した場合にはステップＳ１０１に進み、それ以外の場合にはステップＳ１０２に進む。

［Ｓ１０１］正規化手段８は、日付情報として「from %year-%month-1 to %year-%month-31」を生成する。

［Ｓ１０２］正規化手段８は、%monthの値が４，６，９，または，１１のうちの何れかであるか否かを判定し、ＹＥＳと判定した場合にはステップＳ１０３に進み、それ以外の場合にはステップＳ１０４に進む。

［Ｓ１０３］正規化手段８は、日付情報として「from %year-%month-1 to %year-%month-30」を生成する。

［Ｓ１０４］正規化手段８は、「年」に関する属性値を参照して、閏年か否かを判定し、閏年である場合にはステップＳ１０５に進み、それ以外の場合にはステップＳ１０６に進む。

［Ｓ１０５］正規化手段８は、日付情報として「from %year-%month-1 to %year-%month-29」を生成する。

［Ｓ１０６］正規化手段８は、日付情報として「from %year-%month-1 to %year-%month-28」を生成する。

【００６３】以上の処理によれば、文書に含まれている日付情報が不十分な情報しか含んでいない場合においても、文書の作成日時と、注目する文章の時制とに応じて日付情報を推定するようにしたので、文書に含まれている日付情報を検索を行う際に有効に活用することが可能となる。

【００６４】例えば、「来年の春」のような曖昧な表現も、具体的な数値（例えば、１９９８年３月１日～１９９８年５月３１日）に変換（正規化）することが可能となるので、このような曖昧な表現も検索の際に活用することが可能となる。

【００６５】次に、図１３を参照して、図３のステップＳ１０に示す「金額表現変換処理」の詳細について説明する。このフローチャートが開始されると以下の処理が実行されることになる。

［Ｓ１２０］正規化手段８は、図１４に示す金額表現変換テーブルを参照して、金額表現を対応する数値に変換し、変数ｘに代入する。

【００６６】例えば、「二十万円」という表現では、先ず、「二」が“２”に変換され、「十」が“×10”に、また、「万」が“×10000 ”に変換されるので、全体として値“200000”が得られることになる。

［Ｓ１２１］正規化手段８は、金額表現が「以上」で終わるか否かを判定し、「以上」で終わる場合にはステップＳ１２２に進み、それ以外の場合にはステップＳ１２３に進む。

［Ｓ１２２］正規化手段８は、正規化表現として「from ｘ to ＊」を生成する。ここで、「＊」は任意の値を意味している。

【００６７】前述の例では、x=2000であるので、「from 2000 to *」が生成される。

［Ｓ１２３］正規化手段８は、金額表現が「以下」で終わるか否かを判定し、「以下」で終わる場合にはステップＳ１２４に進み、それ以外の場合にはステップＳ１２５に進む。

［Ｓ１２４］正規化手段８は、正規化表現として「from ＊ to ｘ」を生成する。

［Ｓ１２５］正規化手段８は、金額表現が「台」で終わるか否かを判定し、「台」で終わる場合にはステップＳ１２６に進み、それ以外の場合にはステップＳ１２８に進む。

［Ｓ１２６］正規化手段８は、正規化表現として「from ｘ to ｘ」を生成する。

［Ｓ１２７］正規化手段８は、「to」の後のｘに含まれている“0 ”を“9 ”に全て変換する。

【００６８】例えば、「１０万円台」という表現では、ｘ＝100000となるので、この場合には「to」の後のｘに含まれている“0 ”が“9 ”に全て変換されるので、199999 となる。従って、正規化表現としては、「from 100000 to 199999 」が生成されることになる。

［Ｓ１２８］正規化手段８は、金額表現が「台前半」で

終わるか否かを判定し、「台前半」で終わる場合にはステップＳ１２９に進み、それ以外の場合にはステップＳ１３１に進む。

［Ｓ１２９］正規化手段８は、正規化表現として「from x to x 」を生成する。

［Ｓ１３０］正規化手段８は、「to」の後のｘに含まれている最初の“0 ”を“5 ”に変換する。

【００６９】例えば、「１０万円台前半」という表現では、ｘ＝100000となる。この場合には「to」の後のｘに含まれている最初の“0 ”が“5 ”に変換されるので、150000となる。従って、正規化表現としては、「from 100000 to 150000 」が生成されることになる。

［Ｓ１３１］正規化手段８は、金額表現が「台後半」で終わるか否かを判定し、「台後半」で終わる場合にはステップＳ１３２に進み、それ以外の場合には図３の処理に復帰する。

［Ｓ１３２］正規化手段８は、正規化表現として「from x to x 」を生成する。

［Ｓ１３３］正規化手段８は、「from」の後のｘに含まれている最初の“0 ”を“6 ”に変換する。

［Ｓ１３４］正規化手段８は、「to」の後のｘに含まれている“0 ”を“9 ”に変換する。

【００７０】例えば、「１０万円台後半」という表現では、ｘ＝100000となるので、この場合には「to」の後のｘに含まれている最初の“0 ”が“6 ”にステップＳ１３３において変換され、また、「to」の後のｘに含まれている“0 ”が“9 ”に全て変換されるので、正規化表現としては、「from 160000 to 199999 」が生成されることになる。

【００７１】以上の処理によれば、例えば、漢数字によって記述されている金額表現を対応する数値に変換するとともに、例えば、「以上」や「台前半」などの曖昧な表現を含む金額表現も対応する数値に変換することが可能となる。

【００７２】次に、具体的な例を挙げて以上の実施の形態の動作について説明する。いま、図１５に示す文書が図１に示す文書入力部１から入力されたとする。なお、図１５に示す例文は、新製品の発売に関する文書である。

【００７３】このような文書が文書入力部１から入力されると、事象特定手段４は文書に記述されている事象を、知識情報記憶手段３に記憶されている知識情報を参照して特定する（図３のステップＳ２）。

【００７４】図１５の例では、図５の第４行目～１１行目に記述されている「module main」の中の第１番目の項目（＜会社情報＞は［、］？＜日付＞、＜製品＞を発売した。）に該当することから、この文書に記述されている事象が「新製品の発売」であると判定されることになる。

【００７５】なお、図５に示す知識情報では、事象の定義が「module main 」～「end module」によって囲繞された部分に記述されている。また、事象の定義の中に含まれている、例えば、＜会社情報＞などの属性は、「module」や「def 」などにおいて定義されている。例えば、属性＜会社情報＞は、第１７行目～第２１行目の「module」内に定義されており、その内容は、（＜業種＞、＜会社名＞）、（＜業種２＞&連結語；＜会社名＞）、および、（＜会社名＞）の３種類である。

【００７６】ここで、＜業種＞に関する定義は、第１２行目の「def 」の後に記述されており、（．＊メーカ｜．＊会社｜．＊大手｜．＊開発｜．＊販売｜．＊製造｜．＊業）の中の何れかに該当するものが属性＜業種＞の属性値であるとされる。従って、「パソコンメーカ」や「パソコン大手」などの表現は、＜業種＞の属性値であると判定される。なお、「｜」は「or」を意味している。

【００７７】また、同意語または類義語を含めて定義を行う場合には、第１９行目に示されているように、同意語を含める部分を「&」と「；」の間に挿入する。この例では、「連結語」が同意語または類義語を含む部分となり、その詳細は、第１６行目に定義されており、「連結語」＝（を専門とする｜である｜している｜する｜の）となる。従って、「オフィスオートメーションを専門とする橋本電機」という表現は、会社情報の第２番目の定義（＜業種２＞&連結語；＜会社名＞）に該当することになる。

【００７８】このように、本実施の形態においては、トップダウン的な処理が実行されることから、文脈に応じたパターンマッチングが可能となる。以上のような処理によって事象の種類が特定されると、時制取得手段７は、事象が記述されている文章を取得し、その時制情報を取得する。図１５に示す文書の例では、その時制は過去形（「発売した」）であるので、「過去形」が時制情報として取得される。なお、このようにして取得された時制は、図１６の第２行目に示すように「アスペクト＝過去」として、正規化情報に付加される。

【００７９】次に、属性値抽出手段５は、特定された事象の種類に応じて、属性値を抽出する（図３のステップＳ３）。即ち、属性値抽出手段５は、図５に示す知識情報と文書との間でパターンマッチングを行うことによって属性値を抽出する。

【００８０】図１５の例では、例えば、＜組織名＞として「橋本電機」が抽出され、また、新たに発売する＜製品情報＞の＜種別＞としては「ＪＣＮ互換パソコン」が抽出され、その＜製品名＞としては、「ＧＮＷシリーズ」が抽出されている。

【００８１】続いて、正規化手段８は、文書に日付表現が存在するか否かを判定し（図３ステップＳ４）、存在する場合には対応する数値に変換する処理を行う。図１５に示す文書では、「十八日」という表現が含まれてい

ることから、正規化手段８は、図３に示すステップＳ５において文書作成日時情報と時制情報とを取得して、ステップＳ６において日付表現変換処理を行う。

【００８２】例えば、文書作成日時が「１９９３年１０月１９日」であるとすると、図１６の第６行目に示すように、「発表日付」としてそのタイプが「date」であり、また、その値が「1998-10-18」である情報が正規化情報に付加されることになる。

【００８３】続いて、正規化手段８は、図３のステップＳ７において、金額表現が存在するか否かを判定する。図１５に示す文書では、「十七万八千円」などの表現があることから、ステップＳ８に進み、そこで、規定された通貨単位か否かが判定される。例えば、規定された通貨単位が「円」であるとし、対象となる表現が前述の「十七万八千円」である場合には、ステップＳ１０に進むことになる。

【００８４】なお、「＄１５０」などの表現が含まれている場合には、ステップＳ９において交換レート（１＄＊＝１３０円）に応じて、通貨単位の変換が行われた後、ステップＳ１０に進む。

【００８５】ステップＳ１０では、文字列「十七万八千円」が、値“178000”に変換される。続くステップＳ１１では、他の数値表現が存在するか否かが判定されるが、図１５に示す例の第１番目の文章には、日付表現以外の数値表現は存在しないことから、ステップＳ１３に進む。

【００８６】ステップＳ１３では、対応付け手段１０が固有名が存在するか否かを判定する。図１５の例では、固有名「橋本電機」が存在することから、ステップＳ１４に進む。

【００８７】ステップＳ１４では、対応付け手段１０が知識情報記憶手段３に記憶されている知識情報のうち、橋本電機に対応する情報を取得する。なお、この情報は、例えば、以下のような情報である。

【００８８】

０００１　橋本電機＜会社名＞　０００１１　橋本太郎＜社長名＞
０００１２　岡山県＜所在地＞

取得された「橋本電機」に対する候補が複数存在する場合には、橋本電機に関連付けられて記憶されている他の固有名（橋本太郎、岡山県）などが文書中に含まれていないか判定され、候補が絞り込まれることになる。

【００８９】そして、ステップＳ１４において、絞り込みの結果得られた固有名コード（例えば、０００１）が、正規化情報に付与されることになる（図１６第４行目参照）。

【００９０】ステップＳ１５では、対応付け手段１０は、参照表現が存在するか否かを判定する。図１５に示す例では、参照表現は存在しないから、ＮＯと判定されてステップＳ１８において、生成された正規化情報と文書（または、文書が格納されている場所を示す情報）とを文書記憶手段１１に記憶して処理を終了する。

【００９１】図１７は、他の文書例を示している。また、図１８は、図１７に示す文書を処理して得られた正規化情報の一例を示している。図１８の第３行目に示すように、図１７に示す文書に記述されている事象は、合併情報（field ＝合併情報）であり、その時制は過去（アスペクト＝過去）である。また、「発表した」という表現が第１番目の文章中にないことから、第２行目に示すように、「文末表現＝発表述語なし」とされている。

【００９２】更に、第５行目から第２７行目に示されている「合併主体組織情報」の内容としては、第８行目と第１８行目に示されている北海道大木リフトと、東北海道大木リフトとが合併する主体組織であり、それ以外の行には、これらの組織を補足するための＜合併組織補足情報＞が記載されている。

【００９３】第３４行目以降には、分析の対象となった文以外の残りの文章が記載されている。なお、この例では、図１７の第３行目に「同」という参照表現が含まれているので、図１８の第２３行目に示すように「参照先＝前」という記述が追加され、参照表現「同」が、第１３～第１６行目に示されている「芥川龍太郎（０２５１）」（要素２）であることが示されている。

【００９４】次に、以上のようにして生成された正規化情報を参照して、文書を検索する場合の処理の一例について説明する。図１９は、図１に示すユーザインタフェース部２に表示される入力画面の表示例である。この例では、製品の販売情報が記載された文書を検索の対象としている。即ち、＜製品の販売＞が事象として記述された文書が検索の対象とされる。

【００９５】この例では、第１番目に示すボックス「組織名」に、製品を発売した組織名が入力される。また、第２番目に示すボックス「製品種」には、製品の種類が入力される。更に、ボックス「価格」には製品の価格の範囲が入力される。ボックス「発売日」には、発売された日の範囲が入力される。なお、最下行に表示されているボタン「検索」は、全ての入力が終了し、検索を開始する場合に操作される。

【００９６】図２０は、図１９に示す画面に所定のクエリが入力された場合の入力例を示している。この例では、組織名として「ＡＡＡ」が、また、製品種として「パソコン」が入力されている。

【００９７】更に、価格は、「１００００００」円以上「３０００００」円以下とされており、発売日は「１９９７」年「１」月「１」日から「１９９７」年「６」月「３０」日までとされている。

【００９８】このような入力画面から入力されたクエリ

は、各入力項目の属性を示す情報が付与された後、事象特定手段４、属性値抽出手段５、および、対応付け手段１０を介して、文書抽出手段１２に供給される。なお、付与される情報としては、例えば、「ＡＡＡ」に対してはタグ＜組織名＞が付与される。また、価格はタグ＜価格 type=price unit= 円 value = “ from 100000 to 300000”＞に変換される。更に、発売日は、タグ＜発売日 type=date value=“from 1997-1-1 to 1997-6-30 ”＞に変換される。

【００９９】文書抽出手段１２は、ユーザインタフェース部２から供給されたクエリとタグとに対応する属性値を有する文書を文書記憶手段１１から取得する。即ち、文書記憶手段１１には、元の文書とともに正規化情報が記憶されているので、文書抽出手段１２は、この正規化情報に含まれている属性値と、クエリのタグとを照合することにより、所望の文書を抽出する。

【０１００】このようにして検索された結果は、図示せぬ表示装置に表示出力される。図２１は、検索結果を表示する画面のテンプレートを示している。この例では、検索結果の属性値として「組織名」、「製品種」、「製品名」、「価格」、「発売日」、および、「見出し」が表示される。

【０１０１】図２２は、実際の表示例を示している。この例の第１行目の項目は、「ＡＡＡ」という組織が、デスクトップ型のパソコンを、200000～299999円で、1997/02/29に発売しており、その文書の見出しは「低価格パソコン発売」であることを示している。

【０１０２】図２３は、図１に示すユーザインタフェース部２に表示される入力画面の他の表示例である。この例では、「組織の合併情報」が記載された文書を検索の対象としている。即ち、組織の合併が事象として記述された文書が検索の対象とされる。この例では、第１番目と第２番目に示すボックス「組織名」に、合併する組織名が入力される。また、ボックス「合併日」には、合併が行われる日の範囲が入力される。なお、最下行に表示されているボタン「検索」は、全ての入力が終了した後に、検索を開始する場合に操作される。

【０１０３】図２４は、図２３に示す入力画面に所定のクエリが入力された場合の入力例を示している。この例では、組織名として「ＡＡＡ」が、また、合併日として「１９９７」年「１」月「１」日から、「１９９７」年「１２」月「３１」日までが入力されている。

【０１０４】このような入力画面において、ボタン「検索」が操作されると、前述の場合と同様にタグが生成され、文書記憶手段１１に記憶されている正規化情報と、このタグとを照合することにより、文書が検索される。

【０１０５】図２５は、図２４の検索結果を表示する画面の表示例である。この表示例では、検索結果の属性として「組織名」、「組織名」、「新組織名」、「合併日」、および、「見出し」が表示される。

【０１０６】図２６は、実際の表示例を示している。この例では、検索結果の文書には、組織名が「ＡＡＡ」および「ＢＢＢ」である会社が「1997/04/01」に合併し、新組織名は「ＣＣＣ」であることが示されており、また、その文書の見出しは、「AAA, BBB, ２社合併」であることが示されている。

【０１０７】以上の実施の形態によれば、検索の対象となる事象に対応した入力画面を用意して、その入力画面から必要な項目を入力することにより、所望の文書が取得されることになる。ところで、文書記憶手段１１に記憶されている文書には、前述の正規化情報が関連付けられて記憶されているので、その正規化情報を参照することにより、例えば、対象とする文書に、新たに発売されたパソコンの価格が「二十五万円」と漢数字で記載されているような場合においても、「200000」円～「300000」円と記述されたクエリによって取得されることになる。

【０１０８】なお、以上の実施の形態においては、検索しようとする事象に対応した入力画面から所定の項目を入力し、入力された項目に対応する文書を検索するようにしたが、クエリを文章として入力し、入力された文章に対して正規化処理を行った後、対応する文書を検索するようにしてもよい。以下、そのような方法により、クエリを正規化する処理の一例について、図２７を参照して説明する。このフローチャートが開始されると以下の処理が実行されることになる。

［Ｓ１５１］ユーザインタフェース部２は、文章として記述されたクエリを入力する。

［Ｓ１５２］事象特定手段４は、クエリに記述されている事象の種類を特定する。即ち、事象特定手段４は、知識情報記憶手段３に記憶されている事象と表現とのマッピング規則情報（図５参照）を参照して、クエリに記述されている事象の種類を特定する。

［Ｓ１５３］属性値抽出手段５は、知識情報記憶手段３に記憶されている知識情報を参照して属性値を抽出する。

［Ｓ１５４］正規化手段８は、抽出された属性値に日付表現が含まれているか否かを判定し、日付表現が含まれている場合にはステップＳ１５５に進み、それ以外の場合にはステップＳ１５７に進む。

［Ｓ１５５］作成日時取得手段６は、クエリが作成された日時を取得し、また、時制取得手段７は、クエリの時制（tense ）を取得する。

［Ｓ１５６］正規化手段８は、取得したクエリ作成日時情報と時制情報とを参照して、日付表現を対応する数値に変換する「日付表現変換処理」を実行する。なお、この処理の詳細については、図６を参照して既述したので、その説明は省略する。

［Ｓ１５７］正規化手段８は、抽出された属性値に金額表現が含まれているか否かを判定し、金額表現が含まれ

ている場合には、ステップＳ１５８に進み、それ以外の場合にはステップＳ１６１に進む。

［Ｓ１５８］正規化手段８は、注目している金額表現が規定された通貨単位であるか否かを判定し、規定された通貨単位である場合にはステップＳ１６０に進み、それ以外の場合にはステップＳ１５９に進む。例えば、規定されている通貨単位が「円」ある場合に、「＄」を単位とする金額表現が既述されている場合にはステップＳ１５９に進む。

［Ｓ１５９］単位変換手段９は、内蔵されている記憶部に記憶している交換レートを読み出し、規定されている通貨単位に変換する処理を行う。

【０１０９】例えば、「＄１００」という表現が存在している場合に、交換レートが「１＄＝１３０円」であるとすると、「＄１００」は「１３０００円」に変換されることになる。

［Ｓ１６０］正規化手段８は、金額表現を数値に変換する「金額表現変換処理」を行う。なお、この処理の詳細は、図１３を参照して既述したので、その説明は省略する。

【０１１０】前述の例では、「１３０００円」（文字列）が「13000 」（数値）に変換されることになる。

［Ｓ１６１］正規化手段８は、他の数値表現が存在するか否かを判定し、他の数値表現が存在する場合にはステップＳ１６２に進み、その他の場合にはステップＳ１６３に進む。

【０１１１】例えば、「出荷台数５００００台」などが存在する場合には、ステップＳ１６２に進む。

［Ｓ１６２］正規化手段８は、属性値に含まれている数値表現を対応する数値に変換する。例えば、前述の例では、文字列「５００００」が計算可能な数値「5000」に変換されることになる。

［Ｓ１６３］対応付け手段１０は、属性値に固有名（例えば、「橋本電気」等）が含まれ[illegible]否[illegible]を判定し、含まれている場合にはステ[illegible]進み、それ以外の場合にはステップＳ１[illegible]

［Ｓ１６４］対応付け手段１０は[illegible]有名を抽出し、その固有名に対応する固有名コー[illegible]識情報記憶手段３から取得して属性値に付与する。

【０１１２】例えば、前述の「橋[illegible]気」に対応する固有名コード「００００１１」が、知識情報記憶手段から読み出されて付与される。なお、知識情報記憶手段３には、関連する固有名を相互に関連づけて生成した情報が記憶されているので、ある固有名が複数の候補を有する場合には、関連付けられている他の固有名を参照することにより、固有名を正確に特定することが可能となる。

【０１１３】即ち、「橋本電気」に対応する候補が「橋本電気株式会社」と「株式会社橋本電気」の２つである場合（同名の会社が存在する場合）には、クエリ中に記載されている、例えば、社長名や[illegible]地など[illegible]知識情報記憶手段３に関連付けられて記憶されている固有名とを比較することにより、これらの候補を絞り込んで正確な固有名コードを取得することができる。

［Ｓ１６５］対応付け手段１０は、参照表現（同または両などの表現）が存在するか否かを判定し、参照表現が存在する場合にはステップＳ１６６に進み、それ以外の場合にはステップＳ１６８に進む。

【０１１４】例えば、参照表現である（同）が存在する場合には、ステップＳ１６６に進むことになる。

［Ｓ１６６］対応付け手段１０は、参照表現が参照している対象を同定する。

【０１１５】例えば、「橋本電気（中山社長）は、橋本電算機（同）の独立を・・・」の場合では、参照表現「同」が参照している対象として「中山社長」を同定する。なお、この同定方法としては、「同」や「両」が内挿された括弧“（）”が検出された場合には、その括弧より前に出現する括弧内の属性値を、参照表現が参照している対象であると同定するようにすればよい。

［Ｓ１６７］対応付け手段１０は、参照表現が参照している対象に対応する固有コードを取得し、取得した固有コードを参照表現に対して付与する。

【０１１６】前述の例では、「中山社長」の固有コード「００００１０」が参照表現「同」に付与されることになる。

［Ｓ１６８］対応付け手段１０は、以上のようにして生成されたクエリの正規化情報を、文書抽出手段１２に供給する。その結果、文書抽出手段１２は、以上のようにして生成されたクエリの正規化情報を参照して、文書記憶手段１１に記憶されている文書を検索する。

【０１１７】例えば、クエリとして「橋本酒造が純米酒、橋本を発売した。」が入力された場合には、事象特定手段４は、知識情報記憶手段３に記憶されている知識情報を参照し、入力されたクエリが「新製品の発売」という事象を示していることを特定する。

【０１１８】属性値抽出手段５は、＜組織名＞として「橋本酒造」を抽出し、また、＜製品種＞として「純米酒」を、また、＜製品名＞として「橋本」を抽出する。対応付け手段１０は、「橋本酒造」に対応する固有名コードが存在する場合には、その固有名コードを取得して付与する。例えば、「橋本酒造」の固有名コードが“０１１１”である場合には、＜組織名＞橋本酒造（０１１１）＜／組織名＞というタグが生成されることになる。

【０１１９】文書抽出手段１２は、以上のようにして生成された正規化情報を参照して、文書記憶手段１１から該当する文書を抽出する。即ち、文書抽出手段１２は、組織名タグと固有名コード（０１１１）が付けられた「橋本酒造」、製品種タグが付けられた「橋本」、および、製品名タグが付けられた「橋本」を含み、その事象が「新製品の発売」である文書を文書記憶手段１１から抽出する。

【０１２０】このような処理によれば、例えば、「橋本さんは、橋本酒造製の純米酒を注文した。」という一文が含まれている文書が検索結果として取得されることを防止することができる。即ち、クエリと文書の正規化情報には、抽出された属性を示すタグがそれぞれ付与されていることから、例えば、＜製品名＞である「橋本」を、＜人名＞と混同すること防止することができる。

【０１２１】次に、以上の実施の形態により文書をクリッピングする場合の処理の一例について説明する。図２９は、文書をクリッピングする場合に、ユーザから送信されたクエリを正規化する処理の一例を説明するフローチャートである。このフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ１８０］ユーザインタフェース部２は、所定のユーザからのクエリを入力する。［Ｓ１８１］事象特定手段４、属性値抽出手段５、および、対応付け手段１０は、図２７および図２８のステップＳ１５１～Ｓ１６７の処理を実行して、クエリを正規化する。

［Ｓ１８２］文書抽出手段１２は、正規化されたクエリ（正規化情報）と、それを送信したユーザを特定する情報とを関連付けて記憶する。

［Ｓ１８３］文書抽出手段１２と重要度算出手段１３は、文書記憶手段１１に記憶されている文書と、ユーザ毎のクエリの関連度を判定する「関連度判定処理」を実行する。なお、この処理の詳細は、図３０を参照して後述する。

【０１２２】次に、図３０を参照して、図２９に示す「関連度判定処理」の詳細について説明する。このフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ２０１］重要度算出手段１３は、正規化情報が付与された文書と、正規化されたクエリとの関連度をユーザ単位で計算する。

【０１２３】なお、計算方法としては、例えば、正規化されたクエリに含まれている重要表現を対象となる文書がいくつ含んでいるかに応じてスコアリングを行い、スコアの高い文書を関連度の高い文書とする方法を採用することができる。

［Ｓ２０２］文書抽出手段１２は、重要度算出手段１３の計算結果を参照し、関連度の高い文書を抽出する。

［Ｓ２０３］文書抽出手段１２は、正規化したクエリに含まれている日付、金額、および、数値が、文書の正規化情報に含まれているそれらの値と一致する文書を抽出する。

［Ｓ２０４］文書抽出手段１２は、一致した文書をネットワーク２１を介してユーザに送付する。

【０１２４】続いて、図３１を参照して、例えば、サーバ２３から新たな文書が送信されてきた場合に、文書処理装置２０において実行される処理の一例を説明する。このフローチャートが開始されると、以下の処理が実行されることになる。

［Ｓ２３０］文書入力部１は、ネットワーク２１を介して、例えば、サーバ２３から新たな文書の入力を受ける。

［Ｓ２３１］事象特定手段４、属性値抽出手段５、および、対応付け手段１０は、文書の正規化処理を実行する。

【０１２５】即ち、事象特定手段４、属性値抽出手段５、および、対応付け手段１０は、図３および図４に示す処理を実行することにより入力された文書に対応する正規化情報を生成する。

［Ｓ２３２］文書抽出手段１２および重要度算出手段１３は、図３０に示す「関連度判定処理」を実行する。その結果、生成された正規化情報に一致するクエリが存在している場合には、そのクエリを送信したユーザに対して、新たに入力された文書が送付される。

【０１２６】以上の処理によれば、新たな文書が入力された場合には、入力された文書の正規化情報と、各ユーザの正規化されたクエリとの関連度を算出して、関連度が高い場合には対応するユーザに対して文書を送信するようにしたので、ユーザの要求に適合した文書を正確に選択して送信することが可能となる。

【０１２７】なお、上記の処理機能は、コンピュータによって実現することができる。その場合、文書処理装置が有すべき機能の処理内容は、コンピュータで読み取り可能な記録媒体に記録されたプログラムに記述されており、このプログラムをコンピュータで実行することにより、上記処理がコンピュータで実現される。コンピュータで読み取り可能な記録媒体としては、磁気記録装置や半導体メモリ等がある。

【０１２８】市場に流通させる場合には、ＣＤ－ＲＯＭ(Compact Disk Read Only Memory)やフロッピーディスク等の可搬型記録媒体にプログラムを格納して流通させたり、ネットワークを介して接続されたコンピュータの記憶装置に格納しておき、ネットワークを通じて他のコンピュータに転送することもできる。コンピュータで実行する際には、コンピュータ内のハードディスク装置等にプログラムを格納しておき、メインメモリにロードして実行するようにすればよい。

【０１２９】

【発明の効果】以上説明したように本発明では、対象となる文書に記述されている事象を特定し、特定された事象に関する属性の属性値を抽出し、抽出した属性値と実世界の実体とを対応付けすることによって生成された情報を参照して、文書を検索またはクリッピングするようにしたので、各属性値を正確に認識して文書を検索またはクリッピングすることが可能となるので、結果として、文書の検索またはクリッピング精度を向上させることが可能となる。

【図面の簡単な説明】

【図１】本発明の実施の形態の構成例を示すブロック図である。
【図２】図１に示す文書処理装置を含む通信システムの構成例である。
【図３】文書の正規化処理の一例を説明するフローチャートである。
【図４】文書の正規化処理の一例を説明するフローチャートである。
【図５】知識情報の一例を示す図である。
【図６】図３に示す日付表現変換処理の詳細を説明するフローチャートである。
【図７】数字変換テーブルの一例を示す図である。
【図８】日時表現変換テーブルの一例を示す図である。
【図９】図６に示す日付推定処理の詳細を示すフローチャートである。
【図１０】図９に示す%year 推定処理の詳細を説明するフローチャートである。
【図１１】図９に示す%month推定処理の詳細を説明するフローチャートである。
【図１２】図９に示す%day推定処理の詳細を説明するフローチャートである。
【図１３】図３に示す金額表現変換処理の詳細を説明するフローチャートである。
【図１４】金額表現変換テーブルの一例を示す図である。
【図１５】図１に示す実施の形態に入力される文書の一例である。
【図１６】図１５に示す文書を処理した結果生成される正規化情報の一例である。
【図１７】図１に示す実施の形態に入力される文書の他の一例を示す図である。
【図１８】図１７に示す文書を処理した結果生成される正規化情報の一例である。
【図１９】製品販売情報に関する文書を検索する際の入力画面の一例である。
【図２０】図１９に示す入力画面に入力がなされた場合の一例である。
【図２１】図１９に示す入力画面に対応する検索結果表＊示画面の一例である。
【図２２】図２０に示す入力画面に対応する検索結果の画面の一例である。
【図２３】組織合併情報に関する文書を検索する際の入力画面の一例である。
【図２４】図２３に示す入力画面に入力がなされた場合の一例である。
【図２５】図２３に示す入力画面に対応する検索結果表示画面の一例である。
【図２６】図２４に示す入力画面に対応する検索結果の画面の一例である。
【図２７】クエリに対する正規化処理の一例を説明するフローチャートである。
【図２８】クエリに対する正規化処理の一例を説明するフローチャートである。
【図２９】文書のクリッピングを行う場合において、ユーザからのクエリに対する処理の一例を説明するフローチャートである。
【図３０】図２９に示す関連度判定処理の詳細を説明するフローチャートである。
【図３１】文書のクリッピングを行う場合において実行される、文書に対する処理の一例を説明するフローチャートである。

【符号の説明】
１　文書入力部
２　ユーザインタフェース部
３　知識情報記憶手段
４　事象特定手段
５　属性値抽出手段
６　作成日時取得手段
７　時制取得手段
８　正規化手段
９　単位変換手段
１０　対応付け手段
１１　文書記憶手段
１２　文書抽出手段
１３　重要度算出手段

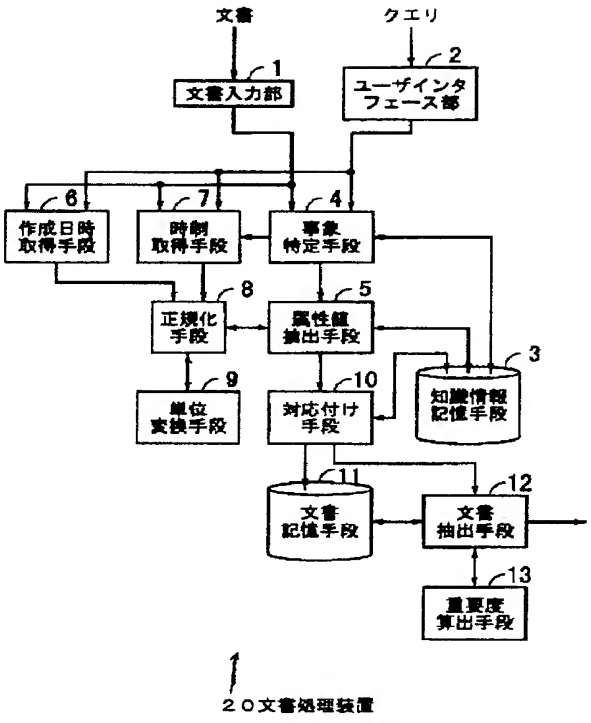
【図１５】

橋本電機は十八日、ＪＣＮ互換パソコン「ＧＮＷシリーズ」を発売したと発表した。これまでは独自設計のパソコンを開発、販売してきたが、昨年からの低価格競争に対抗するため、世界的な業界標準であるＪＣＮ互換機を投入する。橋本電機が発売するのはデスクトップ型とノート型の合計六機種十七モデル。全機種にジョウフムのＣＰＵ（中央演算処理装置）「５９７」とソフィスティケーティッドソフトのＯＳ（基本ソフト）、Ｖ－ＯＳ９８を搭載した。個人ユーザーから企業のネットワーク向けまでを対象としている。台湾メーカーから部品供給を受けるなど海外部品調達をこれまでの平均三〇％から七〇％に高め、生産コストを引き下げた。価格はデスクトップ型が十七万八千円から三十七万八千円まで。ノート型はモノクロディスプレータイプが二十二万八千円から三十四万八千円まで、カラータイプが四十二万八千円から七十四万八千円まで。

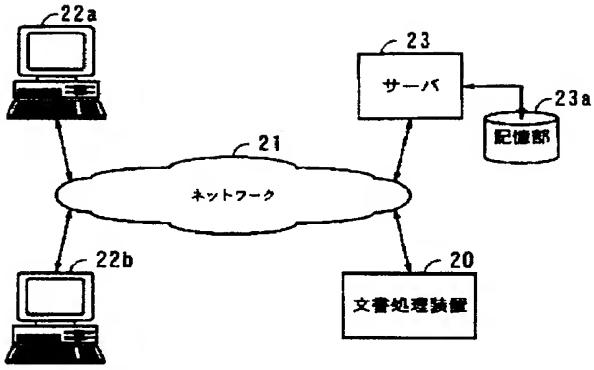
【図１７】

大木リフトの道内販売子会社、北海道大木フォークリフト（石狩管内石狩町、芥川龍太郎社長）と東北海道大木フォークリフト（十勝管内芽室町、同）が一日、合併した。大木リフトが全国で進めている販売強化策に基づくもので、経営を一元化することにより、効率的な営業体制を整える狙い。新会社の名称は北海道大木フォークリフトで、社長には芥川社長が就任した。資本金は二億四千万円、従業員は百三十人。

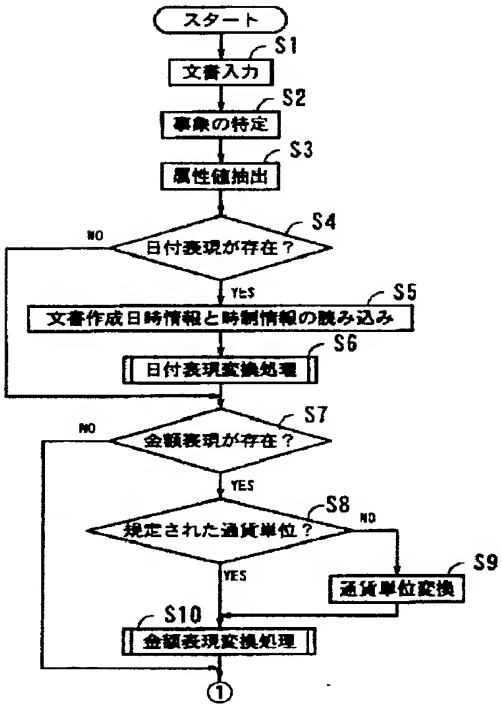
【図１】



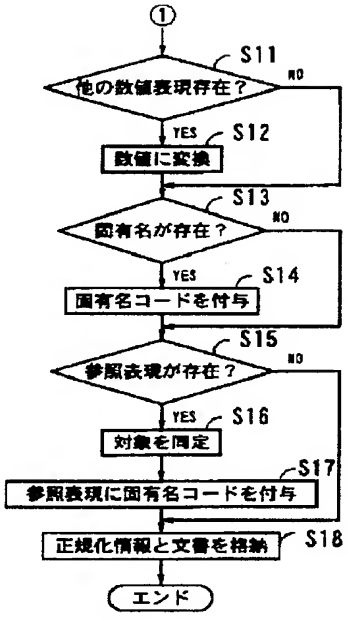
【図２】



【図３】



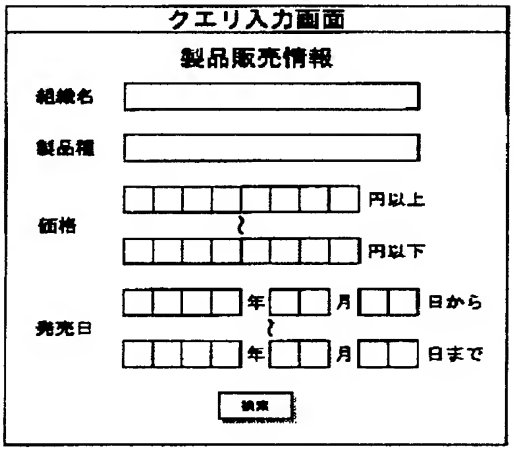
【図４】



【図７】

表現	正規化数値	表現	正規化数値
0	0	7	7
〇	0	七	7
1	1	8	8
一	1	八	8
2	2	9	9
二	2	九	9
3	3	１０	10
三	3	十	10
4	4	１１	11
四	4	十一	11
5	5	１２	12
五	5	十二	12
6	6		
六	6		

【図１９】



【図５】

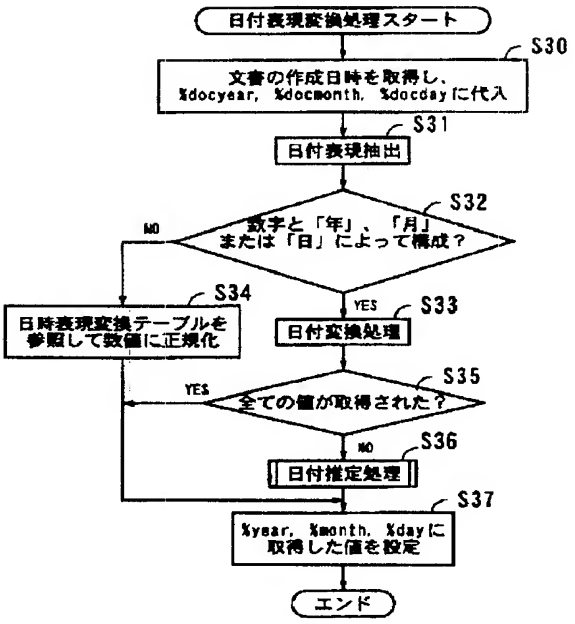
```
typedef 名前(. *)
tagclass 名前<会社情報>, <製品>, <変更点>, <行為>
def <日付> ([一二三四五六七八九十]*日)
module main
  <会社情報>は[、]?<日付>、
                <製品>を発売した。
  <会社情報>は[、]?<製品>を発売した。
  <会社情報>は[、]?<製品>を[、]?<変更点>し[て、]<日付>発売した。
  <会社情報>は[、]?<製品>を<日付>発売した。
  <会社情報>は[、]?<製品>を<行為>、<日付>発売した。
endmodule

  def<業種>  (. *メーカ|. *会社|. *大手|. *開発|. *販売|. *製造|. *業)
  def<業種2> (. *)
  def<会社名> (. *)
  def<会社補足情報> (. *)
synset 連結語 を専門とする である している. する の

module 会社情報
  <業種>, <会社名>
  <業種2>&連結語,<会社>
  <会社名>
endmodule

module 会社名
  <会社名>(<会社補足情報>)
endmodule
```

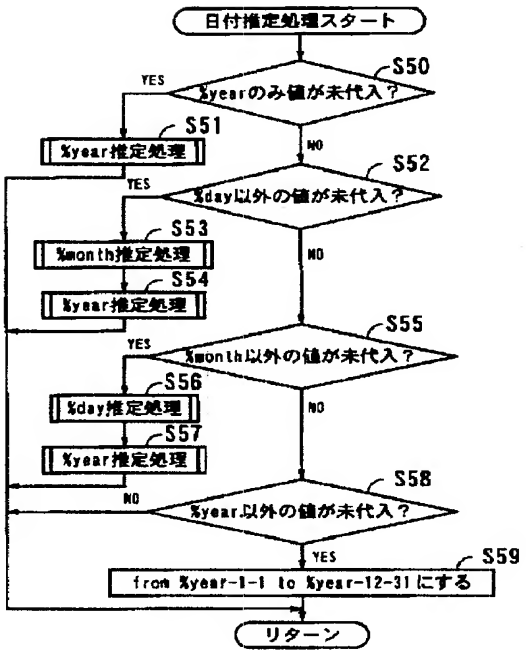
【図６】



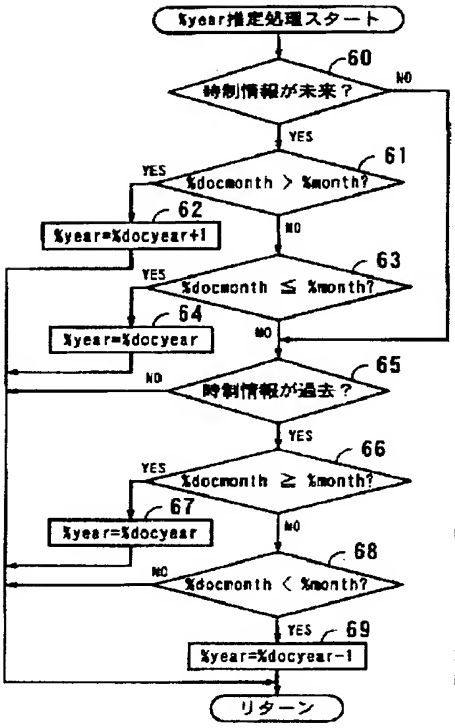
【図８】

表現	type	対応する正規化数値
昨年	date	%docyear - 1
去年	date	%docyear - 1
一昨年	date	%docyear - 2
前年	date	%docyear - 1
来年	date	%docyear + 1
翌年	date	%docyear + 1
本年	date	%docmonth
先月	date	%docmonth - 1
前月	date	%docmonth - 1
来月	date	%docmonth + 1
今月	date	%docmonth
昨日	date	%docday - 1
前日	date	%docday - 1
明日	date	%docday + 1
今日	date	%docday
本日	date	%docday
月初め	daterange	from %year-%month-1 to %year-%month-10
月末	daterange	from %year-%month-21 to %year-%month-31
上旬	daterange	from %year-%month-1 to %year-%month-10
中旬	daterange	from %year-%month-11 to %year-%month-20
下旬	daterange	from %year-%month-21 to %year-%month-31
平成	date	%docyear + 1988
春	daterange	from %year-3-1 to %year-5-31
夏	daterange	from %year-6-1 to %year-8-31
秋	daterange	from %year-9-1 to %year-11-30
冬	daterange	from %year-12-1(o %year-2-29
お正月	daterange	from %????-1-1 to %????-1-7
ゴールデンウィーク	daterange	from %????-4-26 to %????-5-5
GW	daterange	from %????-4-26 to %????-5-5
クリスマス	daterange	from %????-12-1to %????-12-25
入学式シーズン	daterange	from %????-4-1 to %????-4-30
受験シーズン	daterange	from %????-1-10 to %????-3-30
梅雨	daterange	from %????-6-1to %????-7-10

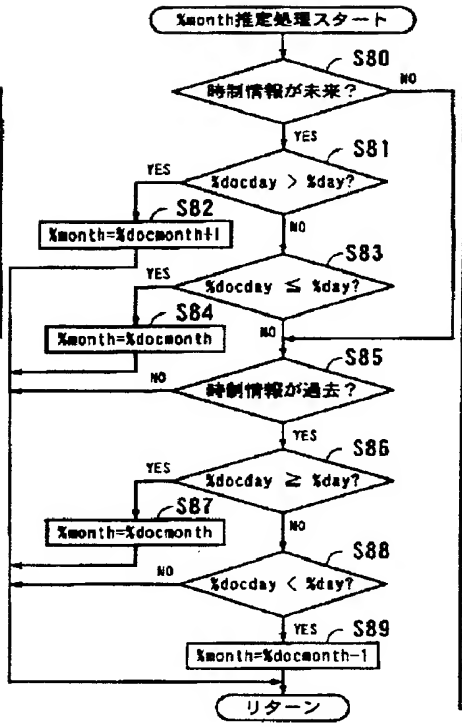
【図９】



【図１０】



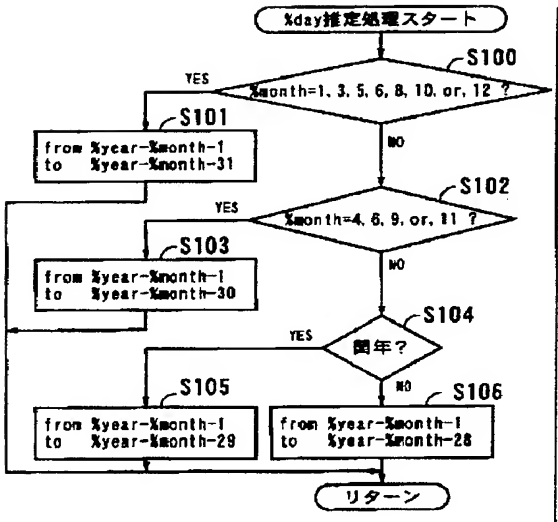
【図１１】



【図１４】

表現	正規化数値	表現	正規化数値
0	0	6	6
○	0	六	6
ゼロ	0	7	7
零	0	七	7
1	1	8	8
一	1	八	8
壱	1	9	9
2	2	九	9
二	2	十	×10
弐	2	百	×100
3	3	千	×1000
三	3	万	×10000
参	3	億	×100000000
4	4	兆	×1000000000000
四	4	京	×10000000000000000
5	5	分	×0.01
五	5	厘	×0.001

【図１２】

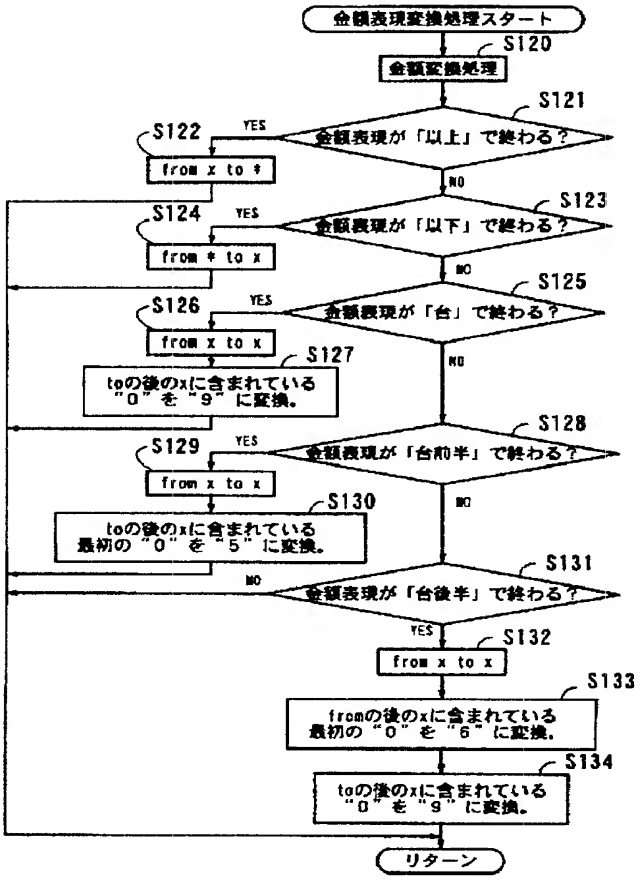


【図１６】

```
<add-anno>
  <記事内容 文末表現=発表述語あり アスペクト=過去>
    <発表主体組織情報>
      <組織名>橋本電機(0001)</組織名>
    </発表主体組織情報>
    <発表日付 type=date value=1993-10-18>
      <日>十八日</日>
    </発表日付>
    <記事内容-用言型1 field=製品販売情報 アスペクト=過去>
      <販売情報-用言型1>
        <販売情報-用言型10>
          <販売製品情報>
            <製品情報>
              <種別>JCN互換パソコン</種別>
              <製品名>GNWシリーズ</製品名>
            </製品情報>
          </販売製品情報>
        </販売情報-用言型10>
      </販売情報-用言型1>
    </記事内容-用言型1>
  </記事内容>

  <記事内容残り>これまでは独自設計のパソコンを開発、販売して きたが、
        昨年からの低価格競争に対抗するため、世界的な業界標準
        であるJCN互換機を投入する。橋本電機が発売するのは
        デスクトップ型とノート型の合計六機種十七モデル。全機
        種にジョウフムのCPU(中央演算処理装置)「697」と
        ソフィスティケーティッドソフトのOS(基本ソフト)
        ソー0398を搭載した。個人ユーザーから企業のネット
        ワーク向けまでを対象としている。台湾メーカーから製品供給
        を受けるなど国外製品調達をこれまでの平均三〇%から七〇%
        に高め、生産コストを引き下げた。価格はデスクトップ型が
  </記事内容残り>
  <価格 type=price unit=円 value="from 178000 to 178000">
    <金額>十七万八千</金額>
  </価格>
  <記事内容残り>から三十七万八千円まで。ノート型はモノクロディスプレ
        ータイプが二十二万八千円
```

【図１３】



【図１８】

```
<add-anno>
 <記事内容 文末表現=発言述語なし>
  <記事内容・用言型O field=合併情報 アスペクト=過去>
   <合併情報・用言型O>
    <合併主体組織情報 type=合併主体関組織情報 >
     <関係組織名>大木リフト(020)</関係組織名>
     <関係組織業種>道内販売子会社</関係組織業種>
     <組織名>北海道大木フォークリフト(021)</組織名>
     <合併組織補足情報1 type=組織補足情報 >
      <要素1 要素数=2 type=要素 >
       <組織所在地>石狩管内石狩町</組織所在地>
      </要素1>
      <要素2 要素数=2 type=要素 >
       <氏名>井川順太郎(0251)</氏名>
       <役職名>社長</役職名>
      </要素2>
     </合併組織補足情報1>
     <組織名>東北海道大木フォークリフト(021)</組織名>
     <合併組織補足情報2 type=組織補足情報 >
      <要素1 要素数=2 type=要素 >
       <組織所在地>十勝管内芽室町</組織所在地>
      </要素1>
      <要素2 要素数=2 type=要素 参照タイプ=同1 参照先=前(0251)>
       <参照表現>同</参照表現>
      </要素2>
     </合併組織補足情報2>
    </合併主体組織情報>
    <合併日付 type=date value=1994-04-01>
     <日>一日</日>
    </合併日付>
   </合併情報・用言型O>
  </記事内容・用言型O>
 </記事内容>
 <記事内容残り>
  <記事内容残りO>大木リフトが全国で進めている販売強化策に基づくもので、経営を一元化することにより、効率的な営業体制を整える狙い。</記事内容残りO>
  <新組織名>北海道大木フォークリフト</新組織名>
  <記事内容残りO>、社長には井川社長が就任した。資本金は</記事内容残りO>
 </記事内容残り>
 <価格 type=price unit=円 value='from 240000000 to 240000000'>
  <資本金>二億四千万</資本金>
 </価格>
 <記事内容残り>、従業員は百三十人。</記事内容残り>
</add-anno>
```

【図２０】

クエリ入力画面

製品販売情報

組織名	AAA
製品種	パソコン
価格	1000000 円以上 ～ 3000000 円以下
発売日	1997年1月1日から ～ 1997年6月30日まで

検索

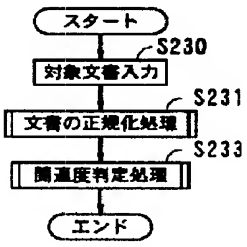
【図２１】

検索結果表示画面

製品販売情報

組織名	製品種	製品名	価格	発売日	見出し

【図３１】



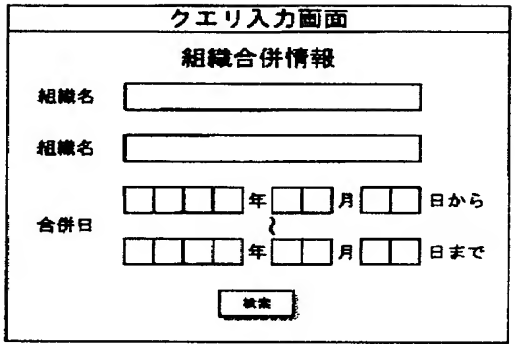
【図２２】

検索結果表示画面

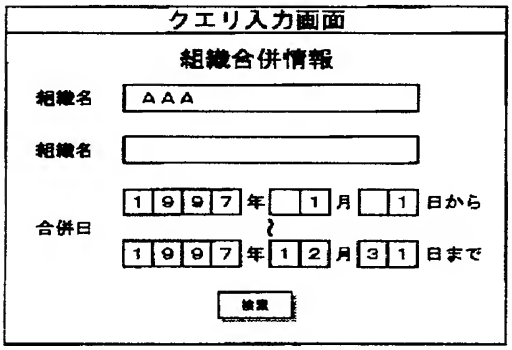
製品販売情報

組織名	製品種	製品名	価格	発売日	見出し
ＡＡＡ	デスクトップ型パソコン		200,000～299,999	1997/02/29	低価格パソコン発売
ＡＡＡ	ディスプレイ一体型パソコン	ＢＢＢＢＢ	268,000	1997/03/01	ＢＢＢＢＢ新シリーズ発売
ＡＡＡ	カラーノートパソコン	ＣＣＣＣＣ	298,000	1997/04/11	ＡＡＡ新発売
ＡＡＡ	互換型パソコン	ＤＤＤＤＤ	178,000	1997/05/20	互換型機発売

【図２３】



【図２４】



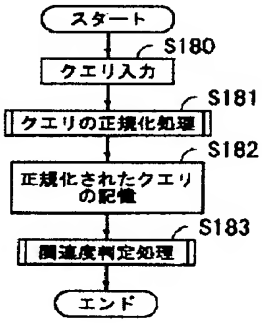
【図２５】

検索結果表示画面

組織合併情報

組織名	組織名	新組織名	合併日	見出し

【図２９】



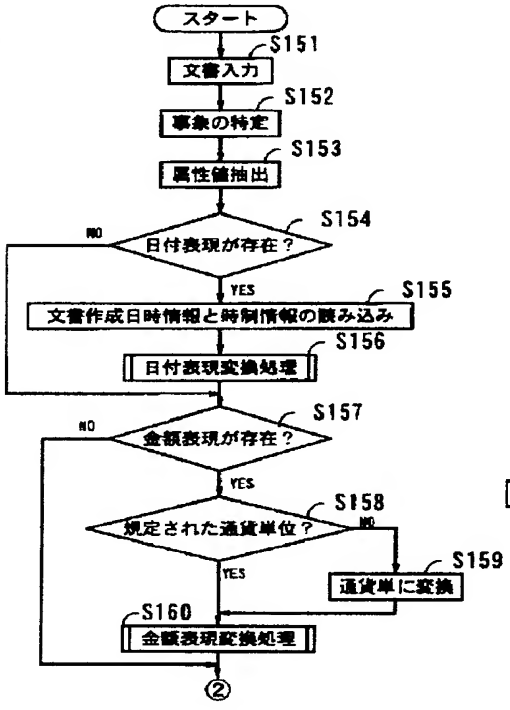
【図２６】

検索結果表示画面

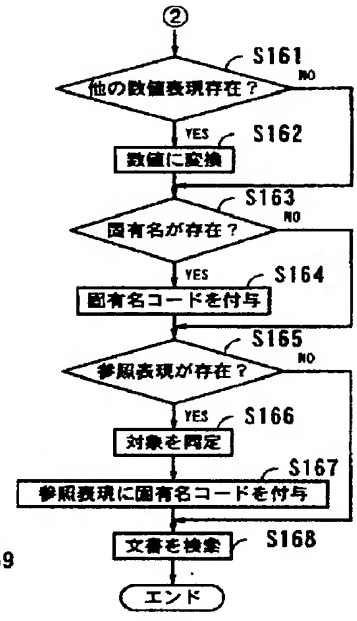
組織合併情報

組織名	組織名	新組織名	合併日	見出し
ＡＡＡ	ＢＢＢ	ＣＣＣ	1997/04/01	AAA, BBB, 2社合併

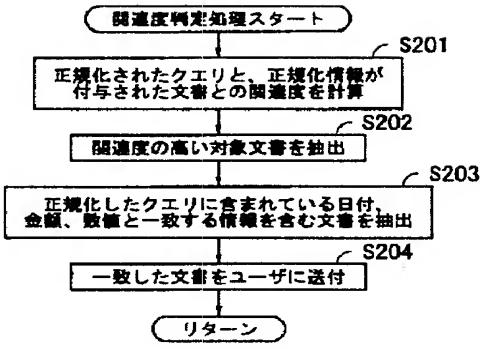
【図２７】



【図２８】



【図３０】



フロントページの続き

(72)発明者　落谷　亮
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

(72)発明者　西野　文人
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内